Time-Varying Perturbations Can Distinguish Among Integrate-to-Threshold Models for Perceptual Decision Making in Reaction Time Tasks

Xiang Zhou
xiangz@math.princeton.edu
KongFatt Wong-Lin
kfwong@math.princeton.edu
Program in Applied and Computational Mathematics, Princeton University,
Princeton, NJ 08544, U.S.A.

Philip Holmes
pholmes@math.princeton.edu
Program in Applied and Computational Mathematics and Department of Mechanical
and Aerospace Engineering, Princeton University, Princeton, NJ 08544, U.S.A.

Several integrate-to-threshold models with differing temporal integration mechanisms have been proposed to describe the accumulation of sensory evidence to a prescribed level prior to motor response in perceptual decision-making tasks. An experiment and simulation studies have shown that the introduction of time-varying perturbations during integration may distinguish among some of these models. Here, we present computer simulations and mathematical proofs that provide more rigorous comparisons among one-dimensional stochastic differential equation models. Using two perturbation protocols and focusing on the resulting changes in the means and standard deviations of decision times, we show that for high signal-to-noise ratios, drift-diffusion models with constant and time-varying drift rates can be distinguished from Ornstein-Uhlenbeck processes, but not necessarily from each other. The protocols can also distinguish stable from unstable Ornstein-Uhlenbeck processes, and we show that a nonlinear integrator can be distinguished from these linear models by changes in standard deviations. The protocols can be implemented in behavioral experiments.

1 Introduction

Reaction time tasks have long been used to study human decision making (Luce, 1986). One paradigm requires subjects to detect or discriminate sensory signals by making a voluntary motor response. Response time

---

Note: Xiang Zhou and KongFatt Wong-Lin contributed equally to this work.

distributions obtained in such perceptual-motor tasks allow inference of cognitive information processing (Posner, 1978). With the advent of methods for recording neural activity in awake behaving animals, such tasks have been adopted by neurophysiologists, especially in studies on monkeys (Schall, 2003; Gold & Shadlen, 2007), and neuronal firing rates in several brain areas have been shown to correlate with motor responses. The lateral intraparietal area, frontal eye fields, and superior colliculus (Gold & Shadlen, 2007) all exhibit activities that ramp up over time toward a fixed level before a decision is signaled (e.g., by a saccadic eye movement in the direction of the recorded response field) (Hanes & Schall, 1996; Roitman & Shadlen, 2002; Churchland, Kiani, & Shadlen, 2008). The slopes of the ramps not only correlate with task difficulty (the harder the task, the lower the slope) but also with response time (higher slopes precede faster responses). These areas may therefore provide neural substrates for integrating sensory information toward a decision criterion before a perceptual decision is made.

Various integrate-to-threshold models have been proposed to describe both response times and neurobiological mechanisms (Smith & Ratcliff, 2004), including drift-diffusion models (Ratcliff, 1978; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Smith & Ratcliff, 2004; Ditterich, 2006a; Simen, Cohen, & Holmes, 2006; Ratcliff & McKoon, 2008) and attractor neural networks with mutual inhibition (Brown & Holmes, 2001; Usher & McClelland, 2001; Wang, 2002; Wong & Wang, 2006; Lo & Wang, 2006). All share a common mechanism: accumulation or integration of sensory inputs toward a prescribed threshold, the first crossing of which determines the decision and response time. These models may nonetheless be distinguished by the details of the integration process: drift-diffusion models typically accumulate evidence at a constant rate like a biased random walk, while attractor networks have unstable or stable steady states that can, respectively, accelerate or decelerate the integration process (Usher & McClelland, 2001; Wong & Wang, 2006).

Often only behavioral data are collected for human subjects, and fits based on response times and choice accuracies are sometimes unable to distinguish among competing models (Ratcliff, Zandt, & McKoon, 1999; Ratcliff, 2006). Moreover, while cellular recordings in awake behaving animals offer direct insights into the integration process, fitting of both behavioral and neural data may still not suffice, especially when models incorporate multiple features and depend on multiple parameters. For example, a drift-diffusion model with a time-dependent urgency signal (Ditterich, 2006a; Churchland et al., 2008) may be difficult to separate from a recurrent network model with strong self-excitation (Wang, 2002; Wong & Wang, 2006). Thus far, few principled attempts have been made to tease apart different integration mechanisms.

Subthreshold electrical microstimulation of neural activities in behaving animals may provide more conclusive tests (Cohen & Newsome, 2004). Ditterich, Mazurek, & Shadlen (2003) showed that microstimulation of

sensory neurons affected the speed of the decision. More interestingly, in Hanks, Ditterich, and Shadlen (2006), stimulation of cells with choice targets in the recorded neuronal response fields hastened decisions, but also reduced decision speeds when choice target directions were opposite to the recorded response fields, thus providing evidence of mutual inhibition. Perturbations need not be invasive, so they can be used in human studies: Huk and Shadlen (2005), for example, employed a brief motion pulse in the background of the primary visual stimulus; although it neither determined the choice nor influenced rewards, the pulse had a significant effect on response times.

Earlier modeling efforts (Huk & Shadlen, 2005; Wong, Huk, Shadlen, & Wang, 2007; Wong & Huk, 2008) have addressed the data of Huk and Shadlen (2005), but they employed many parameters, and direct comparisons were not made between models. Here, we conduct a more rigorous study using simpler models and seek more objective comparisons among them. We approximate four neural firing rate models as linear, scalar, stochastic differential equations (SDE) and ask how their predicted response times are affected by short piecewise-constant perturbations with varying onset times and amplitudes. By comparing changes in means and standard deviations of response times, we demonstrate that the perturbations suffice to distinguish among the models. Finally, we show that a nonlinear integrator model, qualitatively similar to that of Wong and Wang (2006), behaves much like one of the linear models.

2 Methods

2.1 Reduction to One-Dimensional Linear Integrate-to-Threshold Models.
Two-alternative forced-choice decision processes can, in essence, be modeled by two populations of excitatory neurons, each endowed with self-excitatory connections and mutual inhibition via a shared inhibitory population. Each of the three populations can then be represented in a coarsely grained firing rate model by a single unit whose state describes the population-averaged activity $r_j(t)$ of the corresponding neuronal pool (Wilson & Cowan, 1972, 1973). Here we review the further reduction of a firing rate model, under suitable hypotheses, to a one-dimensional dynamical system. (For additional details, see Brown, Gao, Holmes, Bogacz, & Cohen, 2005; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006.)

Consider first the deterministic equations describing the firing rates of two excitatory (E) populations and a common inhibitory (I) population:

$$\tau_E \frac{dr_1}{dt} = -r_1 + F_E(r_1, r_2, r_I, I_1),$$

$$\tau_E \frac{dr_2}{dt} = -r_2 + F_E(r_2, r_1, r_I, I_2), \qquad (2.1)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + F_I(r_1, r_2).$$

Here τ_E and τ_I are the synaptic time constants for the E and I units, and r_1, r_2, and r_I are, respectively, the activity of neural units 1, 2, and I. The excitatory populations 1 and 2 are selective to stimuli 1 and 2, respectively. F_E and F_I are their input-output functions, and the overall input to each unit, with $i = E$ or I, is $I_i = I_{recurrent,i} + I_{stimulus,i}$. The decision time is the first passage time from stimulus onset to the first of r_1 or r_2 reaching a prescribed decision threshold, which thereby signals choice 1 or 2. Since nondecision latencies (e.g., signal transduction and motor preparation) are usually assumed to be independent of stimulus strength $I_{stimulus}$, we model the response time as the decision time plus a constant latency. Henceforth, we use the phrases *decision time*, *reaction time*, and *response time* interchangeably.

If the decision dynamics passes near a saddle point and moves along its unstable manifold (Brown & Holmes, 2001; Bogacz et al., 2006), the functions F_E, F_I may be linearized such that equation 2.1 simplifies to

$$\tau_E \frac{dr_1}{dt} = -r_1 + (\alpha r_1 - \beta r_I + I_1 + I_{0,E}),$$

$$\tau_E \frac{dr_2}{dt} = -r_2 + (\alpha r_2 - \beta r_I + I_2 + I_{0,E}), \tag{2.2}$$

$$\tau_I \frac{dr_I}{dt} = -r_I + (\gamma r_1 + \gamma r_2 + I_{0,I}).$$

Here α, β, γ are the recurrent synaptic coupling strengths for self-excitation, inhibitory-to-excitatory, and excitatory-to-inhibitory connections, and $I_{0,E}$ ($I_{0,I}$) is the constant background input to all the excitatory (inhibitory) cells from outside the local circuit. Unlike Wang (2002), we exclude excitatory connections between r_1 and r_2 and self-inhibitory connections, retaining only essential features. If $\tau_I \ll \tau_E$, we can further assume that the relatively fast dynamics of r_I equilibrates rapidly, such that $r_I \approx \gamma(r_1 + r_2) + I_{0,I}$, and equation 2.2 becomes

$$\tau_E \frac{dr_1}{dt} = -r_1 + (\alpha - \beta\gamma)r_1 - \beta\gamma r_2 + I_1 + (I_{0,E} - \beta I_{0,I}), \tag{2.3}$$

$$\tau_E \frac{dr_2}{dt} = -r_2 + (\alpha - \beta\gamma)r_2 - \beta\gamma r_1 + I_2 + (I_{0,E} - \beta I_{0,I}). \tag{2.4}$$

Defining a new variable $X \equiv r_1 - r_2$ and subtracting equation 2.4 from 2.3, we obtain

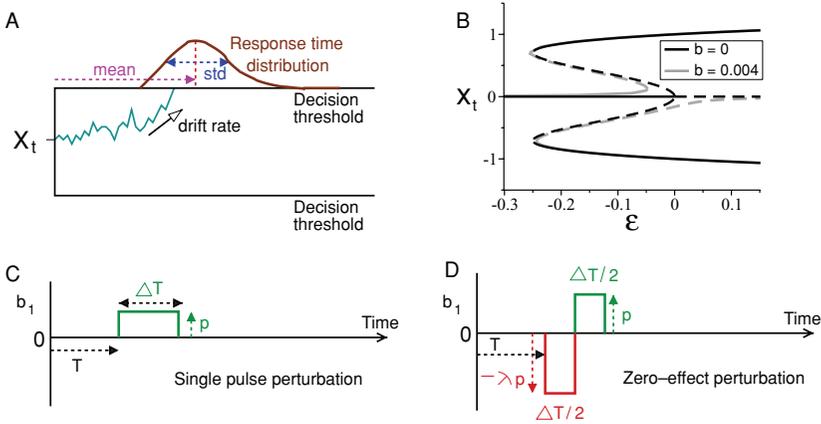$$\frac{dX}{dt} = kX + (I_1(t) - I_2(t))/\tau_E,$$

Figure 1: One-dimensional integrate-to-threshold models and perturbation protocols. (A) A general model for a two-alternative forced-choice reaction time task. Here the noisy state X_t, which quantifies the difference in firing rates between two competing excitatory cell populations, reaches the upper decision threshold in a sample trial, and the choice is correct since threshold is in same direction as drift rate. Multiple sampling provides the response time distribution. (B) The SPB model with unbiased input exhibits a subcritical pitchfork bifurcation at $\varepsilon = 0$. Black curves show branches of stable (solid) and unstable (dashed) fixed points in the noiseless limit; branches for biased input are shown in gray. (C) A short pulse can be repeated over a range of onset times T after primary stimulus onset at $t = 0$ to test effects on decision times. (D) A pulse-antipulse perturbation whose leading pulse amplitude (red) can be adjusted to determine when decision times are unaffected by the combined pulse, yielding a zero-effect perturbation (ZEP). See text for further details.

where $k \equiv (\alpha - 1)/\tau_E$ contains the excitatory coupling strength and leak, and $(I_1 - I_2)/\tau_E$ is proportional to the difference in inputs. The background inputs $I_{0,E}$ and $I_{0,I}$ and coupling strengths β and γ cancel out.

Generalizing the inputs I_1, I_2 to be time varying and including additive noise, the reduced dynamics is described by a one-dimensional SDE of the form

$$dX_t = b(X_t, t)\,dt + \sigma\,dW_t, \tag{2.5}$$

where $b(X_t, t) \equiv kX_t + (I_1(t) - I_2(t))/\tau_E$ and σ is the standard deviation of the noise, which is assumed to be a Wiener process with increments dW_t drawn from a normal distribution with zero mean and unit variance. Equation 2.5 provides a general description for noisy accumulator models, as illustrated schematically in Figure 1A.

2.2 Linear Integrate-to-Threshold Models. We are now in a position to introduce four linear instances of equation 2.5. When $k < 0$ (e.g., self-excitation α is relatively weak), we have a classical Ornstein-Uhlenbeck (OU) process (Uhlenbeck & Ornstein, 1930; Wang & Uhlenbeck, 1945). If the decision threshold does not intervene, solutions eventually decelerate and approach a stable steady-state $X_{steady-state} = (I_1(t) - I_2(t))/(|k|\tau_E)$. We denote this stable OU process SOU; in contrast, if $k > 0$ (stronger self-excitation), sample paths accelerate away from a fixed point in an unstable OU process, denoted UOU. SOU and UOU provide reduced descriptions of the leaky competing accumulator of Usher and McClelland (2001) that respectively produce recency and primacy effects.

In the special case of $\alpha = 1$ (i.e., $k = 0$) and constant stimuli I_1, I_2, equation 2.5 becomes a pure drift-diffusion (CD) equation (Ratcliff, 1978),

$$d X_t = b_0 dt + \sigma d W_t, \tag{2.6}$$

in which the constant drift rate $b = b_0 \equiv (I_1 - I_2)/\tau_E$ is proportional to the difference in stimuli. This CD model is similar to, but differs in detail from, the balance between leakage and inhibition in Bogacz et al. (2006).)

In some extensions of the CD model, drift rates can vary with time. This could be due to dynamic stimuli, to some form of urgency induced by the task design (Ditterich, 2006a, 2006b; Churchland et al., 2008), to fluctuating attention (Smith et al., 2004), or to other top-down effects (Liu, Holmes, & Cohen, 2008). This time-dependent (TD) diffusion model is described by

$$d X_t = b(t) dt + \sigma d W_t. \tag{2.7}$$

We focus on the simple case $b(t) = b_0 t$, since similar results obtain for other functions of time, but in appendix B, we treat a modified TD model with time-varying perturbation amplitude.

2.3 A Nonlinear Integrate-to-Threshold Model. The final model to be considered is qualitatively similar to that of Wang (2002) and Wong and Wang (2006). It captures stochastic dynamics in the neighborhood of a subcritical pitchfork bifurcation (Guckenheimer & Holmes, 1983; Brown & Holmes, 2001; Strogatz, 2001) that occurs as the stimulus input level is varied (Roxin & Ledberg, 2008; Wong & Wang, 2006):

$$\tau_X d X_t = \left[\varepsilon X_t + X_t^3 - X_t^5 + b(t) \right] dt + \sigma d W_t. \tag{2.8}$$

Here $b(t) \equiv b_0$ (constant) and ε represent the biased and nonbiased stimulus inputs respectively (see Figure 1B). τ_X controls the overall temporal dynamics. Equation 2.8 may be derived by normal form theory (Guckenheimer & Holmes, 1983; Roxin & Ledberg, 2008) (see Figure 1B for an illustration of the branches of stable and unstable fixed points in the noise-free limit). If

$|\varepsilon|$ and $|b|$ are sufficiently small, this model without the X_t^5 term approaches those of Wang (2002), Wong and Wang (2006), and Roxin and Ledberg (2008). We include $-X_t^5$ to prevent unrealistic runaway activity and, for simplicity, first assume a symmetrical system with $b_0 = 0$ and then consider biased stimuli $|b_0| > 0$. This model is denoted SPB.

2.4 Perturbation Protocols

2.4.1 Single Pulse Perturbation with Varying Onset Time. We consider additive perturbations $b_1(t)$, under which equation 2.5 becomes $dX_t = [b(X_t, t) + b_1(t)]\,dt + \sigma\,dW_t$, where $b_1(t)$ is applied from time T to $T + \Delta T$ and $b_1(t) \equiv 0$ otherwise. We employ two piecewise-constant forms, the first being the step function used by Huk and Shadlen (2005):

$$b_1(t) = \begin{cases} 0, & t \leq T, \\ p, & T < t \leq T + \Delta T, \\ 0, & t > T + \Delta T, \end{cases} \tag{2.9}$$

in which the amplitude p can be positive or negative, assisting or opposing the unperturbed drift rate b_0 due to the primary stimulus. We fix the duration ΔT at 10% of the unperturbed mean first passage time τ_0, apply b_1 at different onset times T, and ask how the mean and standard deviation of the decision time change.

2.4.2 Zero-Effect Pulse-Antipulse Perturbation. The second protocol uses a double pulse of the form

$$b_1(t) = \begin{cases} 0, & t \leq T, \\ -\lambda p, & T < t \leq T + \Delta T/2, \\ p, & T + \Delta T/2 \leq t < T + \Delta T, \\ 0, & t > T + \Delta T, \end{cases} \tag{2.10}$$

where λ is the relative height of the first pulse to the second. The second opposite-signed pulse attempts to reduce or cancel the expected change due to the first, and if the overall perturbation b_1 leaves the mean first passage time unchanged, we call it a *zero-effect perturbation* (ZEP). This protocol, a variant of the paired pulse suggested in Wong et al. (2007), is applied only early in the integration process to avoid interference from the decision threshold. We shall seek critical values of λ for which ZEPs occur for each model. Figures 1C and 1D illustrate both protocols.

2.5 Simulations and Parameter Values. Except for the SPB model, we first consider integration to a single decision threshold in the direction of

Table 1: Parameters used in Simulations.

Model	$b(X_t, t)$	Parameters	ζ	ΔT	p
CD	$b_0 + b_1$	$b_0 = 5, \sigma = 2.449$	20	0.4	5
TD	$b_0 t + b_1$	$b_0 = 4, \sigma = 2.828$	20	0.1	4
SOU	$k X_t + b_0$ $+ b_1$	$k = -1, b_0 = 8,$ $\sigma = 1.414$	7	0.4	2
UOU	$k X_t + b_0$ $+ b_1$	$k = 0.2, b_0 = 5,$ $\sigma = 1.414$	20	1	2
SPB	$\varepsilon X_t + X_t^3$ $- X_t^5 + b_0$ $+ b_1$	$\varepsilon = -0.3, 0.05, b_0 = 0$ or $0.004, \sigma = 0.01,$ $\tau_X = 20$	± 0.75	5	0.005 or 0.0008 $(b_0 = 0.004)$

Notes: Drift-diffusion model with constant drift rate (CD) and time-varying drift rate (TD); SOU (UOU) stable (unstable) Ornstein-Uhlenbeck processes; SPB: nonlinear model. See equations 2.5 to 2.10 for details. Here ζ is the distance from starting point $X = 0$ to decision threshold, and for SPB $\varepsilon = -0.3$ before the stimulus appears at $t = 0$ and $\varepsilon = 0.05$ for $t \geq 0$. Perturbation durations ΔT and amplitudes p are values used in the first perturbation protocol. Parameters for the linear models with smaller signal-to-noise ratios are specified in section 3.3.

the drift. This reduces the number of parameters, simplifies mathematical analyses, and helps isolate key effects. It applies to easy tasks in which drift rates are (relatively) high and errors rare. We subsequently relax this condition in our simulations. For the SPB model, we first set $I_1 = I_2$, corresponding to difficult tasks and high error rates, and employ two thresholds. We then consider $I_1 \neq I_2$ and a single threshold, representing easy tasks.

Signal-to-noise ratios were chosen such that the variance in first passage (decision) times is significant, but not so great that an unreasonably large number of trials is needed to average out the noise, and we selected perturbation amplitudes, durations, and onset times such that both protocols cause small but significant effects (e.g., maximum changes of approximately 10% in mean decision times). In particular, durations were an order of magnitude smaller than mean decision times τ_0, as in Huk and Shadlen (2005). For the first protocol, onset times T were varied from stimulus onset at $t = 0$ until there were no significant effects on mean decision time. In the second protocol, we require $T \ll \tau_0$. We set $X_0 = 0$ at $t = 0$ to represent unbiased initial conditions, and the remaining parameters are chosen to ensure realistic behavior (e.g., the stable fixed point is above threshold for SOU; the unstable fixed point is below $X_0 = 0$ for UOU). Table 1 lists parameter values.

We use a forward Euler-Maruyama scheme (Higham, 2001), integrating a sample path until it hits the prescribed decision threshold and recording the corresponding first passage time, at which the trial ends and the next begins. After collecting an appropriately large ensemble, we extract the first and second moments of the first passage time from the samples. The simulation is run once for each of the different perturbations. For the models CD, TD,

SOU, and UOU, the step size is 10^{-3} of a time unit and the sample size is $N = 10^6$, so that errors in the moments are of order $1/\sqrt{N} \approx 10^{-3}$. Since changes in moments due to the perturbations are of order 10^{-2}, these choices of time step and sample size reliably capture the effects of perturbation. The mean exit time for the unperturbed SPB model is about 80 time units, so that a time step of 0.01 and a sample size is 10^6 suffices. We ran noisy simulations of this model with $\varepsilon = -0.3$ for $t < 0$ to represent prestimulus activity and then switched to $\varepsilon = 0.05$ for $t \geq 0$ so that $X = 0$ undergoes a pitchfork bifurcation and becomes unstable, thus forcing a choice (see Figure 1B; magnitudes of ε and σ for SPB are much smaller than the corresponding b_0 and σ for the linear models because the $+X_t^3$ term accelerates solutions toward the thresholds).

The parameters are not independent, and their number can be reduced by one in all models by dividing equation 2.5 by b_0 and rescaling $\sigma \to \sigma/b_0$, $p \to p/b_0$, $\zeta \to \zeta/b_0$, $k \to k/b_0$, and the dynamical variable $X_t \to X_t/b_0$. The resulting CD and TD models are described by two parameters, while the SOU and UOU models require three and the SPB model four. Two additional parameters describe the amplitude p and duration ΔT of the perturbation.

We focus on qualitative patterns of changes in response times, and to make unbiased comparisons, we adopt a dimensionless measure of the changes in means and standard deviations by normalizing the relative changes with respect to their unperturbed values.

3 Results

3.1 Single Pulse Perturbations of Linear Models. Using a single pulse perturbation with varying onset time T, we estimated changes in mean decision time for the four linear integrate-to-threshold models as described above, obtaining the results shown in Figure 2. Imposed early during the integration process, a brief pulse in the same (opposite) direction as the drift rate significantly decreases (increases) mean decision times in all four models, as shown by the dashed (solid) curves. These effects fade as T increases due to a thresholding effect: for later perturbations, more trials have already crossed threshold, and so there are relatively fewer trials that are perturbed than unperturbed. Thus, when all trials are averaged, the influence of perturbations is progressively reduced. Note that unlike the work of Huk and Shadlen (2005) and Wong et al. (2007), trials that have crossed threshold before and during perturbation are not excluded in the averaging process. This helps to reduce noise in the data, especially with late perturbation onset times.

In appendix A, we prove that this basic pattern—positive pulses advance mean decision times and negative pulses retard them—must hold provided that the perturbation occurs sufficiently early. The proof applies
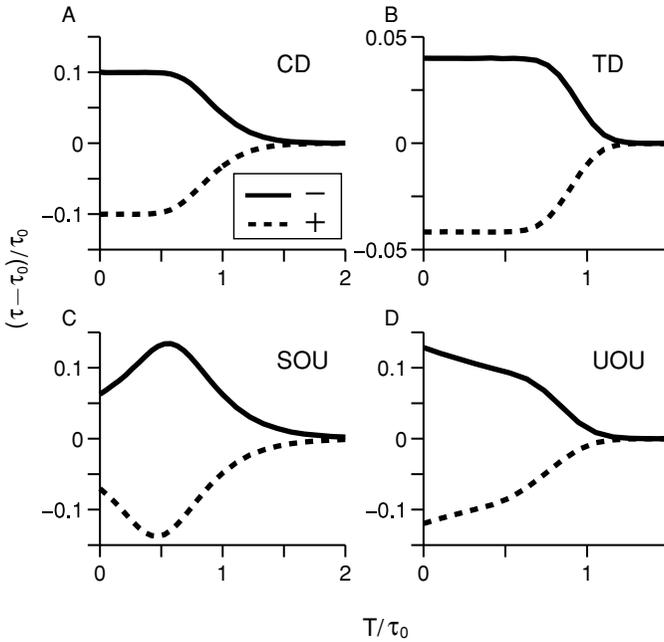
Figure 2: Changes in mean decision times under single-pulse perturbations with varying onset times superimposed on positive drift rates for linear integrators CD, TD, SOU, and UOU (panels A–D). Vertical axes: Normalized changes in mean first passage times $(\tau - \tau_0)/\tau_0$; horizontal axes: normalized perturbation onset times T/τ_0. Changes due to negative pulses shown solid; changes due to positive pulses dashed. CD and TD show near-constant changes for $T/\tau_0 \leq 0.5$ that are substantially greater for CD (note the vertical scale on panel B), SOU exhibits a maximum at $T/\tau_0 \approx 0.5$ (optimal perturbation), and UOU shows a monotonic decrease. Effects decrease in all cases as T approaches and passes τ_0 due to thresholding (see the text).

to a broad class of nonlinear systems including the SPB model of equation 2.8, provided that $\varepsilon > 0$ and thresholds lie inside the region in which drift magnitude increases with X (i.e., well below the activation levels of the stable fixed points).

Note that the drift-diffusion models (CD, TD) and the OU models (SOU, UOU) respond distinctly to early perturbations—the former exhibiting almost constant changes in mean decision times (see Figures 2A and 2B) and the latter respectively showing increasing and decreasing effects (see Figures 2C and 2D). Moreover, due to thresholding, SOU is alone in having an onset time for which the perturbative effect is maximized.

We next investigate how standard deviations are affected by perturbation onset time (see Figure 3). Richer patterns arise than those of Figure 2. Both
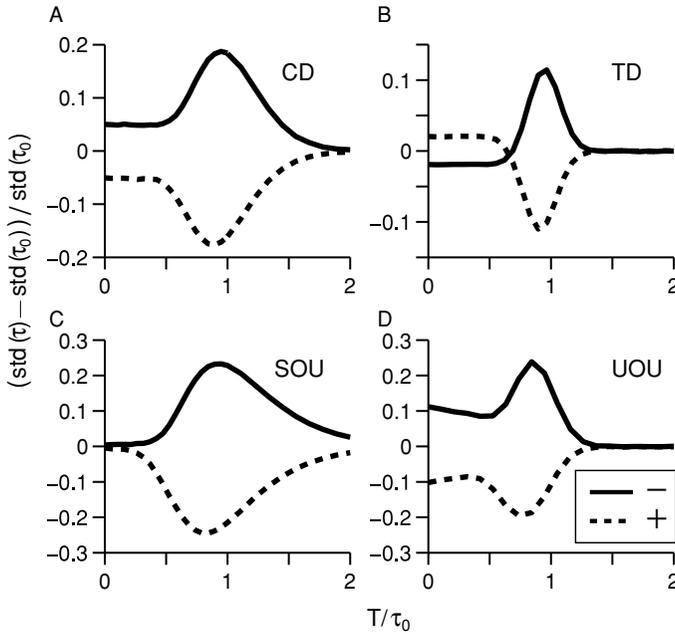
Figure 3: Changes in standard deviations of decision times under single-pulse perturbations with varying onset times. All cases exhibit a maximum in normalized standard deviation $(\text{std}(\tau) - \text{std}(\tau_0))/\text{std}(\tau_0)$, but only UOU has a (shallow) minimum. Effects of early pulses on all four integrators are distinct, as described in text. Format and conventions are as in Figure 2.

CD and TD exhibit similar near-constant changes for early onsets, but the directions of changes for TD are opposite those for CD and the curves cross before reaching maxima. Standard deviations for SOU have a similar pattern to its mean, but UOU exhibits initial decreases followed by increases to a peak, unlike its mean. Optimal onset times exist for which the four standard deviations are maximally affected, and all four cases show distinct patterns.

In the TD model considered above, we assumed that the perturbation is not affected by the time-dependent gain $b(t)$. This may not hold if the perturbation enters by the same sensory pathway as the stimuli. In appendix B, we consider a case in which perturbation and drift are affected in the same way. We show that this modified TD model differs qualitatively only for early perturbation onset, for which it yields reduced effects.

3.2 Zero-Effect Perturbations of Linear Models. We now consider the zero-effect protocol of equation 2.10, appealing to the fact that in the noise-free case, the first passage time is a monotone function of the ratio λ of the

amplitudes of the opposing pulses. The critical $\lambda = \lambda^*$ that determines a ZEP is therefore given by the unique zero of

$$\tau(\lambda) = \tau_0, \tag{3.1}$$

where $\tau(\lambda)$ describes the functional dependence of passage time on λ. Mean first passage times appear to remain monotone with respect to λ when noise is included. Asymptotic methods can be used to approximate passage time moments for small perturbations (Lindner, 2004), but the lack of compact formulas for passage time distributions (PDFs) in most cases, including OU processes, and the need for double integrations (Lindner, 2004) make this approach generally intractable.

However, ZEPs may be analytically approximated for early onset times T with $T' := T + \Delta T \ll \tau_0$, such that the probability of threshold crossing prior to T' is negligible. In this case, a sufficient condition is that the PDFs of the perturbed and unperturbed processes are identical at T', since for $t > T'$, both processes are governed by the same drift and noise. They are therefore indistinguishable in the limit $T, \Delta T \to 0$. Before deriving explicit expressions for λ^* under this assumption, we present the results of numerical simulations with an intuitive explanation.

Since the unperturbed SDE (see equation 2.5) is independent of the current state X_t in the CD and TD cases, we expect that antisymmetric pulses with $\lambda^* = 1$, in which the pulses precisely cancel, will produce ZEPs. In contrast, responses to inputs decay with time for SOU, implying that the first pulse must be larger than the second for their net effect to cancel at perturbation offset $t = T'$ ($\lambda^* > 1$). For UOU, the reverse should hold ($\lambda^* < 1$). Figure 4 confirms that this is the case. Table 2 lists the parameters used in the simulations; note that the early onset time conditions are only weakly satisfied (T' ranges from 25% to 40% of τ_0).

To derive explicit approximations for λ^*, we use a comparison method similar to that of appendix A. Let X_t denote the unperturbed process and Z_t the perturbed process. If $X_0 = Z_0$ at $t = 0$, then $X_T = Z_T$ at the onset time T. When the perturbation ends at $t = T'$, the solution of the unperturbed SDE of equation 2.5 with constant drift rate b_0 is

$$X_{T'} = (X_T + b_0/k)e^{k\Delta T} - b_0/k + \sigma \int_T^{T+\Delta T} e^{k(T'-s)} \, dW_s, \tag{3.2}$$

while the perturbed system satisfies

$$Z_{T'} = (X_T + b_0/k)e^{k\Delta T} - b_0/k + \int_T^{T+\Delta T} e^{k(\Delta T-s)}b_1(s) \, ds$$

$$+ \sigma \int_T^{T+\Delta T} e^{k(T'-s)} \, dW_s. \tag{3.3}$$
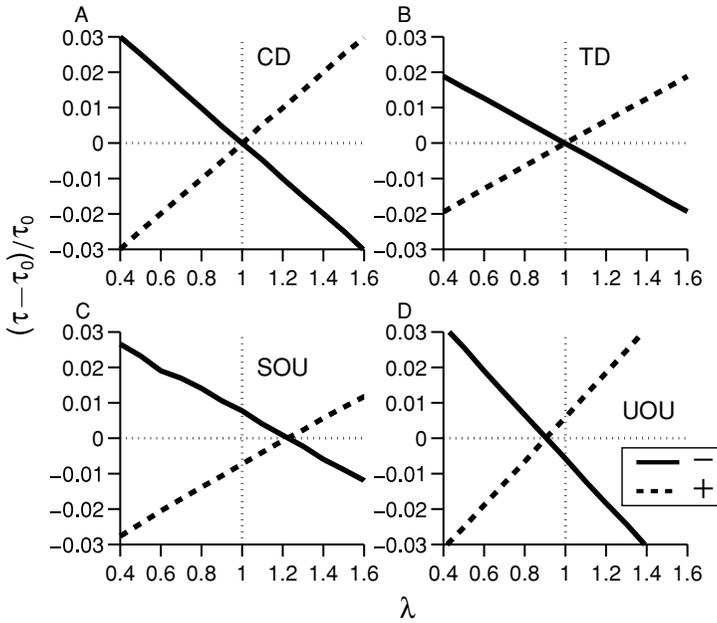
Figure 4: Zero-effect perturbations of linear integrate-to-threshold models CD, TD, SOU, and UOU (panels A–D). Vertical axes: normalized change in mean first passage times $(\tau - \tau_0)/\tau_0$; horizontal axes: relative amplitude λ of first pulse compared to second. Zeroes of solid lines ($-$, positive pulse followed by negative pulse, $p < 0$) and dashed lines ($+$, negative pulse followed by positive pulse, $p > 0$) identify ZEPs at $\lambda^* \approx 0.9987, 1.0039$ (CD); $0.9952, 1.0004$ (TD); $1.2271, 1.2230$ (SOU); and $0.9066, 0.9034$ (UOU).

Table 2: Parameters Used in ZEP Simulations.

Model	τ_0	std(τ_0)	T	ΔT	p	λ^* (thy.)	λ^* (sim.)
CD	4.005	0.960	0.5	0.5	5	1	1.0013
TD	3.145	0.420	0.5	0.5	5	1	0.9978
SOU	1.831	0.374	0.1	0.4	2	1.2214	1.2251
UOU	2.954	0.142	0.2	1.0	2	0.9048	0.9050

Note: τ_0 and std(τ_0) are the first and second moments of the passage time without perturbation; T and ΔT are the onset time and total duration of the perturbation; p is the amplitude of the second pulse; and λ^* (thy.) and λ^* (sim.) are the ZEP pulse ratios predicted by the theory and obtained by averaging $p < 0$ and $p > 0$ results of Figure 4.

These expressions well approximate the true activity levels for models CD, SOU, and UOU with absorbing thresholds only when both processes have a low probability of hitting the threshold during the interval $[0, T']$, but this condition holds for $T'/\tau_0 \ll 1$, since the integrated drift term is almost zero

at early times. Equations 3.2 and 3.3 differ in the term $e^{k\Delta T} \int_0^{\Delta T} e^{-kt} b_1(t) \, dt$, which enters $Z_{T'}$ due to the action of the perturbation b_1 during the interval $[T, T']$. The ZEP condition for arbitrary perturbations is therefore

$$\int_T^{T+\Delta T} e^{-kt} b_1(t) \, dt = 0, \tag{3.4}$$

which for the piecewise constant pulses of equation 2.10 implies that

$$\lambda^* = \frac{e^{-k\Delta T/2} - e^{-k\Delta T}}{1 - e^{-k\Delta T/2}} = e^{-k\Delta T/2}. \tag{3.5}$$

Hence, for $k > 0$ (UOU), $\lambda^* < 1$, while for $k < 0$ (SOU), $\lambda^* > 1$. In the special case of CD, $k = 0$ and $\lambda^* = 1$.

For TD, the drift term is time dependent, but the linear term $k X_t$ is absent, and so the corresponding solutions of the SDEs are

$$X_{T'} = \int_T^{T+\Delta T} b_0(s) \, ds + \sigma \int_T^{T+\Delta T} dW_s$$

and

$$Z_{T'} = \int_T^{T+\Delta T} b_0(s) \, ds + \int_T^{T+\Delta T} b_1(s) \, ds + \sigma \int_T^{T+\Delta T} dW_s.$$

In this case, a ZEP must satisfy

$$\int_T^{T+\Delta T} b_1(t) \, dt = 0, \tag{3.6}$$

which yields $\lambda^* = 1$ for the pulses of equation 2.10, as for CD. This result holds for all time-dependent drift rates, including those used in Smith, Ratcliff, & Wolfgang (2004), Ditterich (2006b), and Churchland et al. (2008).

Summarizing, ZEPs occur for the CD and TD models when the opposing pulse amplitudes are equal ($\lambda^* = 1$), but for SOU and UOU, $\lambda^* = e^{-k\Delta T/2}$ is larger and smaller than one, respectively. The ratios λ^* predicted by equation 3.5 for the parameters of Table 2 are given in the final column of the table. In all four cases, they agree well with the zero crossings of the mean first passage times in Figure 4, in spite of the fact that $T'/\tau_0 \approx 0.25 - 0.4$ is not very small.

3.3 Signal-to-Noise Ratio Can Influence Perturbation Effects. To what extent do the results of section 3.1 hold when signal-to-noise ratios are reduced and a second threshold is added to track errors? To investigate
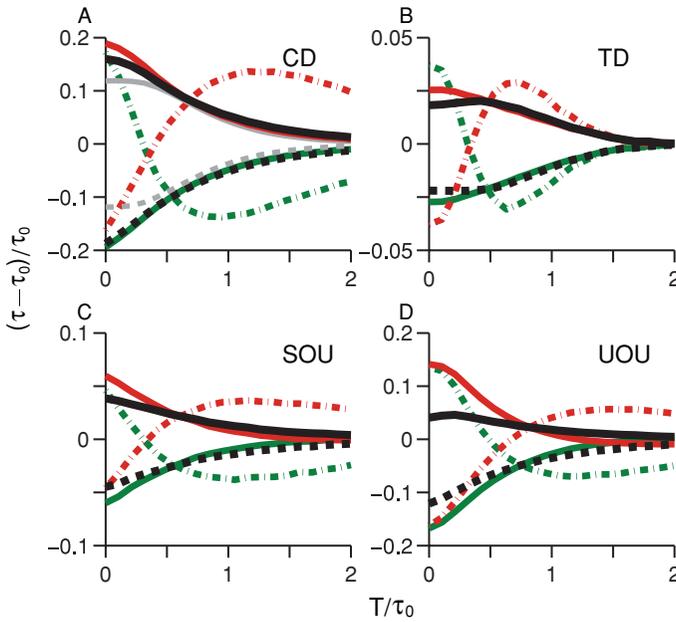
Figure 5: Changes in mean decision times for linear models with lower signal-to-noise ratios. Perturbation protocol is as in Figure 2. Green (red) curves: perturbation p in the same (opposite) direction as drift rate b_0; solid: correct choices; dash-dotted: error choices. Solid (dashed) black curves are averages of correct and error choices with perturbation in the opposite (same) direction as drift rate. Gray (black) curves in panel A are for error rate of 0.02% (5%).

this question, we select new parameter values according to the criteria of section 2.5.

For the CD model, we increase σ from 2.449 to 2.828 and set thresholds at $\zeta = \pm 5$ instead of ± 20, yielding an error rate of 5%. This introduces more complex behavior in which correct and error trials respond to the perturbations in different manners, as shown in Figure 5A. The combined averages (as computed in Huk & Shadlen, 2005, and Wong et al., 2007; see the black solid and dashed curves in Figure 5) preserve some features of previous results: for early perturbations, the direction of changes in mean RT agrees with Figure 2A, but the approximately constant change in mean decision time early in the trial is replaced by a monotonic decline. Increased noise yields earlier threshold crossings, advancing the thresholding effect and masking the signatures of Figure 2A. Moreover, the change in mean error RTs can reverse sign for late perturbation onsets. We checked this seemingly counterintuitive phenomenon numerically by fixing the random generating seed for noise in a sample trial, finding that perturbation of long RT trials can change an impending error to a correct choice (data not shown).

Increasing the noise level to $\sigma = 7.071$ with thresholds $\zeta = \pm20$ produces an error rate of 10% in the TD model and changes in mean RTs for correct and error choices similar to those of CD. However, averaged over correct and error choices, overall changes in mean RT exhibit a pattern similar to that of Figure 2B (see the black curves in Figure 5B). In this respect, TD maintains its signature more robustly with increasing error rate than CD, although the effects are progressively masked.

For the SOU model, we increase σ to 6.325, yielding an error rate of 17%. The optimal perturbation onset time for changes in mean RT averaged over correct and error trials is masked (compare Figure 2C with the black curves in Figure 5C). This error rate is relatively high, but we have checked that for error rates as small as 3%, SOU model features can also be masked, indicating that like the CD model, SOU is sensitive to noise.

For the UOU model, more parameters must be adjusted to obtain reasonable RT changes. Here we set $\sigma = 2$, $\zeta = \pm10$, $k = 0.02$, and $b_0 = 0.5$, yielding a high error rate of 23.5%, but similar trends occur for error rates from 0.1% to 35%. Figure 5D shows that these are comparable to the previous UOU results with high signal-to-noise ratio. Overall, we are unable to distinguish between noise-masking effects and intrinsic features of the UOU integrator.

Using the same parameters, we find that changes in standard deviations for all the linear models show optimal perturbation onset times (see Figure 6). The CD, SOU, and UOU models are mutually indistinguishable (cf. Figures 6A, 6C, and 6D), but TD exhibits a crossing effect (see Figure 6B), similar to Figure 3B.

3.4 Perturbations of the Nonlinear Model. We next apply both perturbation protocols to the nonlinear SPB model. The mean and standard deviation of decision times for the unperturbed system are 80 and 22 time units, respectively. The parameters used for the first protocol are given in Table 1, while for the ZEP, we set $p = 0.005$, $T = 30$, and $\Delta T = 5$, consistent with application early in the accumulation process. Figures 7A and 7B show that the means and standard deviations of decision times behave like those for UOU (cf. Figures 2D and 3D), decreasing monotonically and approaching zero as onset times increase. ZEPs occur in the pulse ratio range $0.8 < \lambda^* < 1$ for both mean and standard deviation (see Figures 7C and 7D). Assuming a linear change in mean in Figure 7C, we find that $\lambda^* \approx 0.87$.

To compare with the behavior of the linear systems in the single threshold case, we now modify the SPB model so that it makes as few errors as they do. This is done by changing b_0 from 0 to 0.004, which breaks the symmetry of the subcritical pitchfork bifurcation (see Figure 1B) so that decisions favor one choice, representing easy tasks (cf. Roxin & Ledberg, 2008); all other parameters remain the same. Figure 8 shows that the basic behavior of Figures 7C and 7D is preserved: effects on both mean and standard deviation of RTs decrease monotonically as perturbation onset time increases. While
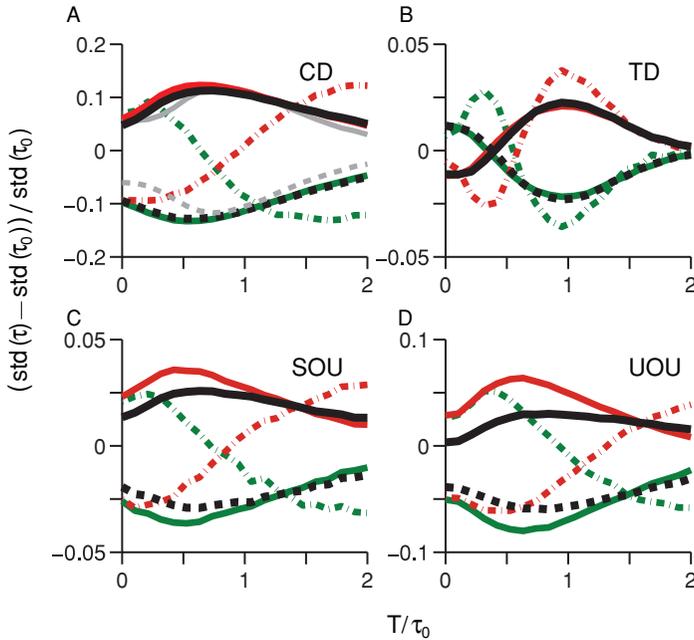
Figure 6: Changes in standard deviations of decision times for the linear models with lower signal-to-noise ratios. Format is as in Figures 3 and 5.

the effects on mean passage times of SPB are much like those for UOU, the fact that standard deviations of all four linear models exhibit maxima (see Figure 3) potentially allows us to distinguish SPB from them. The effects of ZEPs in this biased regime are similar to those shown in Figures 7C and 7D (data not shown).

4 Discussion

In this work, we have investigated how several simple stochastic integrate-to-threshold decision-making models respond to short perturbations. We focus on two-alternative forced-choice reaction time tasks for which general firing rate models with three coupled neural populations can be reduced to one-dimensional integrate-to-threshold systems. These reduced systems, which may be pure drift diffusion or Ornstein-Uhlenbeck processes, can be compared analytically and in simulations that require few parameters.

   We examined two perturbation protocols, the first being a brief pulse with variable onset time and the second a pulse-antipulse pair whose amplitude ratio can be adjusted to produce minimal effects (the zero-effect perturbation, ZEP). The simulations of section 3 (see Figures 2 and 3) show
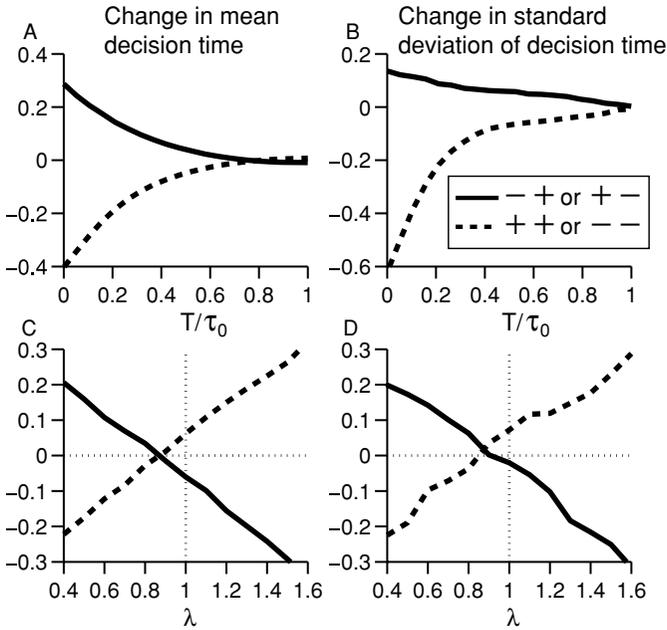
Figure 7: Perturbations of means (left column) and standard deviations (right column) of first passage times for the SPB model. (A, B) Effects of single-pulse perturbations with varying onset times; vertical axes as in Figures 2 and 3. $-+$ denotes the effect of a negative pulse at positive threshold, $+-$ denotes the effect of a positive pulse at negative threshold; and so forth. Due to symmetry, $-+$ and $+-$ have the same effect (solid curves) as do $++$ and $--$ (dashed curves). (C, D) Responses to zero-effect perturbations; vertical axis in C as in Figure 4; vertical axis in D shows normalized change in standard deviation as in Figure 3.

that the changes in means and standard deviations of decision times allow both perturbation protocols to distinguish between drift-diffusion models and stable and unstable Ornstein-Uhlenbeck processes, provided that perturbations are delivered sufficiently early in the integration process and the signal-to-noise ratio is not too small. Changes in standard deviations can also assist in distinguishing drift-diffusion models with constant drift from those with time-dependent drift rates (cf. Figures 3A and 3B), although differences are small and may not be detectable experimentally. (As Ditterich, 2006a, notes, entire reaction time distributions can further assist in this.) For early perturbations, the ZEP conditions can be approximated analytically, making predictions that are confirmed by computer simulations (see Figure 4). While general analytical expressions for systems with finite noise appear elusive, in appendix A we prove inequalities for more general nonlinear systems that partially explain our simulation results.
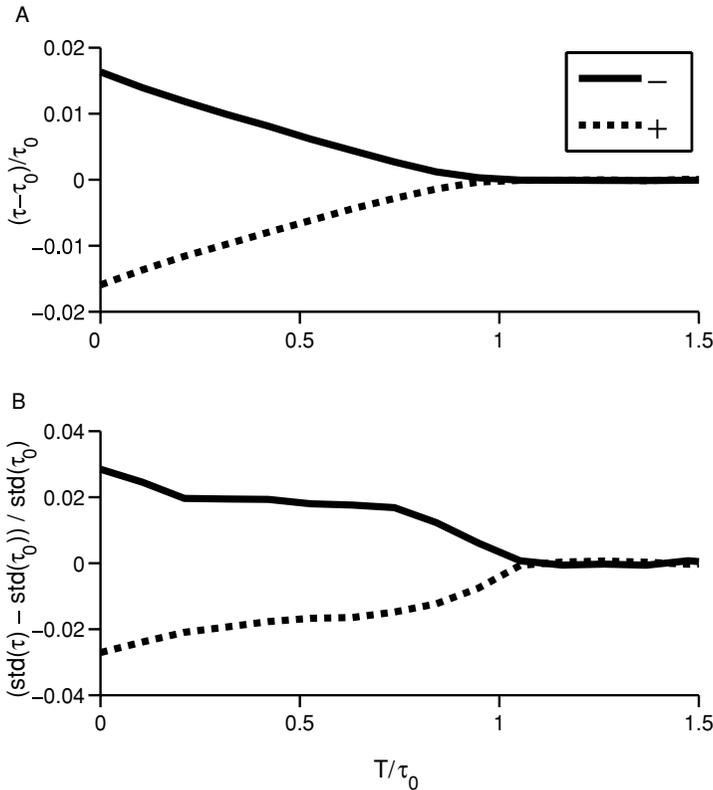
Figure 8: Changes in mean (A) and standard deviation (B) of decision times of the SPB model with high signal-to-noise ratio. All choices are correct. Perturbation protocols and plot formats are as in Figures 7A and B.

Although we focus on linear models, our methods extend to other systems, and we include results for a nonlinear model that capture the dynamics near a subcritical pitchfork bifurcation, which is typical of reduced population models (Wang, 2002; Wong & Wang, 2006; Roxin & Ledberg, 2008). As the results of appendix A suggest, it behaves in a manner similar to the unstable OU model under both perturbation protocols, although perturbative effects on standard deviations may allow distinctions to be made (see Figures 7 and 8). We have also confirmed that the TD model's behavior remains similar to other increasing drift rates (e.g., $b(t) \sim b_0 t^2$; simulation results not shown).

These results reinforce the simulations and claims in Wong et al. (2007) that time-varying perturbations can in principle reveal integration mechanisms. However, some parameter ranges frustrate clear-cut comparisons among models. The first occurs for low signal-to-noise ratios and high

error rates (cf. Figures 5 and 6), and so can be mitigated by using perturbations with more easily discriminated stimuli. Second, if there is a dead time between stimulus onset and the start of evidence integration, the CD and TD models can display similar signatures to the SOU model, even with high signal-to-noise ratios.

Since perturbations can be delivered through the senses as well as by direct electrical stimulation, the method can be noninvasive and is therefore appropriate for human subjects. Furthermore, unlike fitting neuronal firing rates, both perturbation protocols can distinguish linear and nonlinear UOU-type models (Wang, 2002; Wong & Wang, 2006; Roxin & Ledberg, 2008; Wong & Huk, 2008) from drift-diffusion models with an urgent time-dependent increase in drift rate (Smith et al., 2004; Ditterich, 2006a; Churchland et al., 2008). However, care is required in designing such experiments. Although a high signal-to-noise ratio can in principle help identify integrators, the reaction times can become too short, masking the signature behaviors of the integrators with the thresholding effect. A possible solution could be to prolong the response deadline in relatively simple reaction time tasks and analyze only data from long RT trials.

Perturbation inputs to decision-making circuits may originate in a single sensory pathway or in other brain areas, as in multisensory and top-down inputs (e.g., due to attention; Smith et al., 2004; Liu et al., 2008), which could modulate a decision throughout a sensorimotor pathway. The general notion of perturbation developed in the reduced model context extends to such modulatory inputs. The methods may also generalize to other tasks that recruit neural integrators (Goldman, Compte, & Wang, 2009), such as interval timing (Shea-Brown, Rinzel, Rakitin, & Malapani, 2006).

Appendix A: Rigorous Estimates on First Passage Times

In this appendix, we analyze short single-pulse perturbations, first in the noiseless limit and then with additive noise. The inequalities proved here suggest explanations for the monotonicity of mean first passage times of SOU and UOU processes with respect to early onset times (cf. Figures 2C and 2D); moreover, they apply to more general, nonlinear one-dimensional systems such as SPB with thresholds $|\zeta|$ sufficiently close to $X_t = 0$.

A.1 The Noiseless Limit. As noise amplitude tends to zero, the SDE, equation 2.5, becomes an ordinary differential equation (ODE):

$$\frac{dX}{dt} := \dot{X} = f(X). \tag{A.1}$$

Table 3: Monotonicity of First Passage Dependence on Onset Time for Four Sign Combinations of f' and g in the Noise-Free Limit.

	$f' \geq L > 0$	$f' \leq -L < 0$
$g > 0$	$\tau < \tilde{\tau}$	$\tau > \tilde{\tau}$
$g < 0$	$\tau > \tilde{\tau}$	$\tau < \tilde{\tau}$

Note: L is some constant.

Consider two compact perturbations p and $\tilde{p}$ with the same amplitude profile $g(t)$ and duration ΔT, applied at different onset times T_1 and $\tilde{T}_1$:

$$p(t) = 1_{[T_1, T_2]}(t)g(t - T_1), \quad \tilde{p}(t) = 1_{[\tilde{T}_1, \tilde{T}_2]}(t)g(t - \tilde{T}_1) \tag{A.2}$$

(here $1_{[T_1, T_2]}$ denotes the indicator function that takes the value 1 for $T_1 < t < T_2$ and zero otherwise). We choose $0 < T_1 < T_2 < \tilde{T}_1 < \tilde{T}_2$ and $T_2 - T_1 = \tilde{T}_2 - \tilde{T}_1 = \Delta T$, such that p and $\tilde{p}$ do not overlap and that neither perturbed solution reaches the threshold $X = \zeta$ before $t = \tilde{T}_2$. We shall compare solutions Y_t and $\tilde{Y}_t$ of the corresponding perturbed ODEs,

$$\dot{Y} = f(Y) + p(t) \quad \text{and} \quad \dot{\tilde{Y}} = f(\tilde{Y}) + \tilde{p}(t), \tag{A.3}$$

started at the same initial condition $Y_0 = \tilde{Y}_0 = 0$, to show how the signs of f' and g determine their threshold passage times τ and $\tilde{\tau}$. We allow smooth nonlinear functions f but assume that $f(X) > 0$ for all $0 \leq X < \zeta$ and that g does not change sign. The fact that solutions of scalar ODEs cannot cross each other is a key tool.

Lemma 1. *Let $X_t(x)$ denote the solution of equation A.1 with initial value x. Then $X_t(x_1) < X_t(x_2)$ for any $t > 0$ if and only if $x_1 < x_2$.*

Proof. This follows from the uniqueness of solutions of ODEs and the ordering of points in a one-dimensional phase space.

Since neither perturbation acts after $t = \tilde{T}_2$ and neither solution has reached threshold at $\tilde{T}_2$, lemma 1 implies that $\tau < \tilde{\tau}$ if and only if $Y(\tilde{T}_2) > \tilde{Y}(\tilde{T}_2)$. We now prove the main result summarized in Table 3, which is a corollary of proposition 1:

Proposition 1. *Assume that the function $g(t)$ in equations A.2 is strictly positive for $t \in (0, \Delta T)$. Then for some constant L, (a) if $f' \geq L > 0$, then $Y(t) > \tilde{Y}(t)$ for all $t \geq \tilde{T}_2$; (b) if $f' \leq -L < 0$, then $Y(t) < \tilde{Y}(t)$ for all $t \geq \tilde{T}_2$.*

Proof. Part a: We need only establish the inequalities at $t = \tilde{T}_2$, since no perturbations occur after this time. We compare solutions of equations A.3

using the fact that $f(X) > 0$ and $f'(X) \geq L > 0$ for $X \geq 0$ implies that $f(Y) - f(\tilde{Y}) \geq L(Y - \tilde{Y})$. Letting $u(t) = Y(t) - \tilde{Y}(t)$, we have

$$\dot{u} = f(Y) - f(\tilde{Y}) + p(t) - \tilde{p}(t) \geq Lu + p(t) - \tilde{p}(t). \tag{A.4}$$

Since the solutions coincide until the first perturbation occurs at $t = T_1$, $u(T_1) = 0$, and integrating inequality A.4 over the interval $[T_1, T_2]$, on which $p(t) = g(t - T_1)$ and $\tilde{p}(t) \equiv 0$, yields

$$u(T_2) \geq \int_{T_1}^{T_2} e^{L(T_2-t)} g(t - T_1) \, dt > 0. \tag{A.5}$$

Both $p(t)$ and $\tilde{p}(t) \equiv 0$ on $[T_2, \tilde{T}_1]$, and integration of equation A.4 and use of equation A.5 gives

$$u(\tilde{T}_1) \geq u(T_2) e^{L(\tilde{T}_1-T_2)} \geq e^{L(\tilde{T}_1-T_2)} \int_{T_1}^{T_2} e^{L(T_2-t)} g(t - T_1) \, dt > 0. \tag{A.6}$$

Finally, using the fact that $p(t) \equiv 0$ on $[\tilde{T}_1, \tilde{T}_2]$, we integrate equation A.4 again to obtain

$$u(\tilde{T}_2) \geq u(\tilde{T}_1) e^{L(\tilde{T}_2-\tilde{T}_1)} - \int_{\tilde{T}_1}^{\tilde{T}_2} e^{L(\tilde{T}_2-t)} g(t - \tilde{T}_1) \, dt$$

$$\geq \left[e^{L(\tilde{T}_2-T_2)} - 1 \right] \int_0^{\Delta T} e^{L(\Delta T-t)} g(t) \, dt > 0, \tag{A.7}$$

where we also use inequality A.6 and the fact that the additive perturbations of equation A.2 are identical over the time intervals $[T_1, T_2]$ and $[\tilde{T}_1, \tilde{T}_2]$ to rewrite the integral term. Hence, $Y(\tilde{T}_2) > \tilde{Y}(\tilde{T}_2)$ and $\tau < \tilde{\tau}$.

Part b: We again compare solutions of equations A.3, but the direction of the inequalities is now reversed, so that we are in essence using Gronwall's inequality (Guckenheimer & Holmes, 1983). In this case, equation A.4 becomes

$$\dot{u} \leq -Lu + p(t) - \tilde{p}(t); \tag{A.8}$$

inequality A.6 becomes

$$u(\tilde{T}_1) \leq u(T_2) e^{-L(\tilde{T}_1-T_2)} \leq e^{-L(\tilde{T}_1-T_2)} \int_{T_1}^{T_2} e^{-L(T_2-t)} g(t - T_1) \, dt < 0; \tag{A.9}$$

and the final inequality reads

$$u(\tilde{T}_2) = Y(\tilde{T}_2) - \tilde{Y}(\tilde{T}_2) \le -\left[1 - e^{-L(\tilde{T}_2 - T_2)}\right] \int_0^{\Delta T} e^{-L(\Delta T - t)} g(t)\, dt < 0.$$

(A.10)

This implies that $\tau > \tilde{\tau}$, as claimed. Proofs are similar for $g < 0$.

A.2 Extension to Noisy Systems. The conclusions of proposition 1 and Table 3 may be extended to apply to expected passage times for SDEs with additive noise. Specifically, if equations A.3 are replaced by

$$dY_t = [f(Y_t) + p(t)]\, dt + \sigma d W_t,$$ (A.11a)

$$d\tilde{Y}_t = [f(\tilde{Y}_t) + \tilde{p}(t)]\, dt + \sigma d\tilde{W}_t,$$ (A.11b)

we have the following result:

Proposition 2. *Assume that the function $g(t)$ in equations A.2 is strictly positive for $t \in (0, \Delta T)$. Then for some constant L, (a) if $f' \ge L > 0$, then $\mathbb{E}[\tau] < \mathbb{E}[\tilde{\tau}]$; (b) if $f' \le -L < 0$, then $\mathbb{E}[\tau] > \mathbb{E}[\tilde{\tau}]$, where τ and $\tilde{\tau}$ denote the first passage times for equations A.11a and A.11b, respectively.*

Proof. We first compare solutions Y_t and $\tilde{Y}_t$ obtained from equations A.11a and A.11b under the same sample noise path with increments $d\tilde{W}_t \equiv dW_t$. Since the stochastic integral $\int_0^t dW_t$ is almost surely continuous (Feller, 1957) and the integrated noise terms cancel, the comparison of solutions using $u(t) = Y_t - \tilde{Y}_t$ proceeds as in the deterministic case above, and we conclude that $\tau < \tilde{\tau}$ for $g > 0$ and $f' \ge L > 0$, and $\tau > \tilde{\tau}$ for $g > 0$ and $f' \le -L < 0$. Finally, averaging each process over an ensemble of sample paths, we may conclude that the mean first passage times of the two processes satisfy similar inequalities to those in the top row of table 3: $\mathbb{E}[\tau] < \mathbb{E}[\tilde{\tau}]$ and $\mathbb{E}[\tau] > \mathbb{E}[\tilde{\tau}]$ respectively. A similar argument applies to the case $g < 0$.

Appendix B: A TD Model with Time-Varying Perturbation Amplitude

In the body of this letter, perturbations of the TD model were assumed unaffected by the time-dependent gain. Here we extend the simulations to a simple case in which perturbation and stimulus (drift) share the same overall linear increase in amplitude. Specifically, we assume the dynamics to be governed by

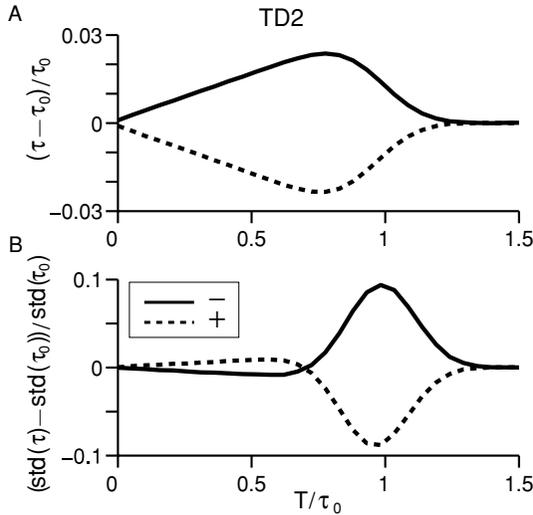$$d X_t = (b_0 + p(t))t\, dt + \sigma d W_t.$$ (B.1)

Figure 9: Normalized changes in mean (A) and standard deviation (B) of first passage times for single-pulse perturbations with varying onset times applied to a variant of the TD model. Vertical axes and curve conventions are as in Figures 2 and 3.

Otherwise, we change only the perturbation amplitude, taking $p = \pm 2$ instead of 4. Figure 9A shows that the changes in mean and standard deviations of exit time for this model remain similar to the previous TD model, but with smaller perturbative effects near the beginning of the trial (quite similar to an SOU model) due to the monotonic time-dependence of perturbation amplitude. However, although the previous pattern of mean exit times no longer applies, the behavior remains qualitatively dissimilar to the CD, UOU, and SPB models.

Acknowledgments

either expressed or implied, of the Air Force Research Laboratory or the U.S. government.

References

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.*, *113*, 700–765.

Brown, E., Gao, J., Holmes, P., R. Bogacz, M. G., & Cohen, J. D. (2005). Simple neural networks that optimize decisions. *Int. J. Bifurcation and Chaos*, *15*, 803–826.

Brown, E., & Holmes, P. (2001). Modeling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics*, *1*, 159–191.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nat. Neurosci.*, *11*, 693–702.

Cohen, M. R., & Newsome, W. T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.*, *14*, 169–177.

Ditterich, J. (2006a). Evidence for time-variant decision making. *Eur. J. Neurosci.*, *24*, 3628–3641.

Ditterich, J. (2006b). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Netw.*, *19*, 981–1012.

Ditterich, J., Mazurek, M. E., & Shadlen, M. N. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nat. Neurosci., 6*, 891–898.

Feller, W. (1957). *An introduction to probability theory and its applications* (2nd ed.). New York: Wiley.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.

Goldman, M., Compte, A., & Wang, X.-J. (2009). Neural integrator models. In L. Squire (Ed.), *Encyclopedia of neuroscience*. Orlando, FL: Academic Press.

Guckenheimer, J., & Holmes, P. (1983). *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*. New York: Springer-Verlag.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*, 427–430.

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat. Neurosci.*, *9*, 682–689.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, *43*, 525–546.

Huk, A. C., & Shadlen, M. N. (2005). A neural integrator underlying perceptual decision-making in macaque parietal cortex. *J. Neurosci.*, *25*, 10420–10436.

Lindner, B. (2004). Moments of the first passage time under external driving. *J. Stat. Phys.*, *117*, 703–737.

Liu, Y. S., Holmes, P., & Cohen, J. D. (2008). A neural network model of the Eriksen task: Reduction, analysis, and data fitting. *Neural Comput.*, *20*, 345–373.

Lo, C.-C., & Wang, X.-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat. Neurosci.*, *9*, 956–963.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cereb Cortex.*, *13*, 1257–1269.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.*, *85*, 59–108.

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognit. Psychol.*, *53*, 195–237.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.*, *20*, 873–922.

Ratcliff, R., Zandt, T. V., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychol. Rev.*, *106*, 261–300.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.*, *22*, 9475–9489.

Roxin, A., & Ledberg, A. (2008). Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Comput. Biol.*, *4*, e1000046.

Schall, J. D. (2003). Neural correlates of decision processes: Neural and mental chronometry. *Curr. Opin. Neurobiol.*, *13*, 182–186.

Shea-Brown, E., Rinzel, J., Rakitin, B. C., & Malapani, C. (2006). A firing rate model of Parkinsonian deficits in interval timing. *Brain Res., 1070*, 189–201.

Simen, P., Cohen, J. D., & Holmes, P. (2006). Rapid decision threshold modulation by reward rate in a neural network. *Neural Netw.*, *19*, 1013–1026.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosci.*, *27*, 161–168.

Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Res.*, *44*, 1297–1320.

Strogatz, S. H. (2001). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering.* Cambridge, MA: Perseus Books Group.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of Brownian motion. *Phys. Rev.*, *36*, 823–841.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychol Rev.*, *108*, 550–592.

Wang, M. C., & Uhlenbeck, G. E. (1945). On the theory of Brownian motion II. *Rev. Mod. Phys., 17*, 323–342.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*, 955–968.

Wilson, H., & Cowan, J. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys., J, 12*, 1–24.

Wilson, H., & Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*, 55–80.

Wong, K.-F., & Huk, A. C. (2008). Temporal dynamics underlying perceptual decision making: Insights from the interplay between an attractor model and parietal neurophysiology. *Front. Neurosci.*, *2*, 245–254.

Wong, K.-F., Huk, A. C., Shadlen, M. N., & Wang, X.-J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision-making. *Front. Comput. Neurosci.*, 1:6:doi:10.3389/neuro.10/006.2007.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.*, *26*, 1314–1328.